LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

LLNL-JRNL-563299

The Transpose-AMIP II experiment and itsapplication to the understanding of Southern Ocean cloud biases in climate models

K. D. Williams, A. Bodas-Salcedo, M. Deque, S. Fermepin, B. Medeiros, M. Watanabe, C. Jakob, S. A. Klein, C. A. Senior, D. L. Williamson

July 6, 2012

Journal of Climate

Disclaimer

The Transpose-AMIP II experiment and its application to the understanding of Southern Ocean cloud biases in climate models

K.D. Williams*[1], A. Bodas-Salcedo[1], M. Déqué[2],

S. Fermepin[3], B. Medeiros[4], M. Watanabe[5],

C. Jakob[6], S.A. Klein[7], C.A. Senior[1] and D.L. Williamson[4]

[1] Met Office, UK

[2] Météo-France/CNRM, CNRS/GAME, France

[3] Institut Pierre Simon Laplace, France

[4] National Center for Atmospheric Research, USA

[5] University of Tokyo, Japan

[6] Monash University, Australia

[7] Lawrence Livermore National Laboratory, USA

*Corresponding author address: Keith Williams, Met Office, FitzRoy Road, Exeter, EX1 3PB, UK.

Email: keith.williams@metoffice.gov.uk Tel: +44 (0)1392 886905 Fax: +44 (0)1392 885681

5th July 2012

Abstract

"Transpose-AMIP" is an international model intercomparison project in which climate models are run in 'weather forecast mode'. The Transpose-AMIP II experiment is running alongside the Coupled Model Intercomparison Project phase 5 (CMIP5) and allows processes operating in climate models to be evaluated, and the origin of climatological biases explored, by examining the evolution of the model from a state in which the large-scale dynamics, temperature and humidity structures are constrained through use of common analyses.

The Transpose-AMIP II experimental design is presented. The project requests participants to submit a comprehensive set of diagnostics to enable detailed investigation of the models to be performed. An example of the type of analysis which may be undertaken using these diagnostics is illustrated through a study of the development of cloud biases over the Southern Ocean, a region which is problematic for many models. Several models share a climatological bias for too little reflected shortwave radiation from cloud across the region. This is found to mainly occur behind cold fronts/on the leading side of transient ridges and be associated with more stable lower tropospheric profiles. Investigation of a case study which is typical of the bias, and associated meteorological conditions, reveals the models to typically simulate cloud which is too optically and physically thin with an inversion which is too low. The evolution of the models within the first few hours suggests these conditions are particularly sensitive and a positive feedback can develop between the thinning of the cloud layer and boundary layer structure. In at least one model, the lack of sufficient vertical resolution to properly represent the boundary layer structure appears to be a factor.

1 Introduction

Over recent years there has been growing interest in using general circulation models (GCMs) across a range of timescales in order to understand the origin of key model biases (e.g. Phillips *et al*, 2004; Rodwell and Palmer, 2007; Martin *et al*, 2010). Running 'weather forecasts' (or more correctly hindcasts since they are run retrospectively) with the atmospheric component of climate models enables detailed evaluation of the processes operating through a comparison of the model with a variety of observations for particular meteorological events (e.g. Boyle and Klein, 2010). In addition, understanding the development of biases as they grow from a well initialized state can provide significant insight into the origin of these biases, which can be used in the future development of the model (e.g. Williamson *et al*, 2005). Many of the principal sources of model spread in terms of simulating climate and climate change are fast-processes (e.g. clouds), hence examining climate models on these timescales can yield greater understanding of why their longer timescale response differs (e.g. Williams and Brooks, 2008; Xie *et al*, 2012).

For those GCMs which are routinely used for both weather and climate prediction, such analysis is commonplace (Senior *et al*, 2010). Indeed the benefits are leading these centers to unify the model science in their prediction systems to a greater extent than ever before (Brown *et al*, 2012). For those climate centers without their own data assimilation system, it has been suggested that many of the benefits of this type of analysis may be realized if they are initialized from an analysis produced by an operational numerical weather prediction (NWP) center. In the USA this was the basis of the Climate Change Prediction Program – Atmospheric Radiation Measurement (CCPP-ARM) Parametrization Testbed (CAPT) in which short range hindcasts from the NCAR (National Center for Atmospheric Research) climate model were initialized from European Centre for Medium

Range Weather Forecasts (ECMWF) analyses (Phillips *et al*, 2004). (After the realignment of some programs, the CAPT acronym was changed to stand for Cloud Associated Parametrization Testbed).

In 2005, the Joint Scientific Committee/Commission on Atmospheric Sciences (JSC/CAS) Working Group on Numerical Experimentation (WGNE) initiated an intercomparison of climate models running a set of short-range hindcasts and initialized from ECMWF analyses. Since WGNE originally set up the Atmospheric Model Intercomparison Project (AMIP; Gates *et al*, 1999) in which atmosphere GCMs were run freely for decades, these short hindcasts in which the large-scale dynamics are constrained by being initialized to a common analysis were termed "Transpose-AMIP" experiments. This set of Transpose-AMIP simulations was undertaken by a small number of centers, and a relatively limited set of diagnostics were collected, with analysis focused on the Southern Great Plains ARM site. Unfortunately given the the mix of phenomena that occurred during the period considered, the forecast sample size was too small to allow a statistically meaningful analysis. In 2008 WGNE agreed that the experiment should be extended and ideally run alongside the CMIP5 (Coupled Model Intercomparison Project version 5; Taylor *et al*, 2012) activity organized by the Working Group on Coupled Modelling (WGCM). The Transpose-AMIP II experiment was subsequently drawn up and became jointly endorsed by WGNE and WGCM.

In this paper we describe the Transpose-AMIP II experiment (hereafter abbreviated to T-AMIP2). To date, data from five GCMs have been submitted to T-AMIP2, with more expected to be submitted over the coming years. We use T-AMIP2 data from these models to illustrate how the experiment, and the range of diagnostics collected, may be used to understand the origin of a bias common to a number of GCMs. Trenberth and

Fasullo (2010) show that many GCMs are biased to have too little reflected shortwave radiation over the Southern Ocean which they argue affects both the coupled model performance due to excess shortwave radiation reaching the ocean, and the reliability of climate change projections due to a possible spurious cloud response over the region. More recently, detailed studies into the climatological structure and distribution of clouds over the Southern Ocean, and their evaluation in a particular GCM, have been carried out by Haynes *et al* (2011) and Bodas-Salcedo *et al* (2012). Here, we relate the climatological Southern Ocean radiation biases seen in the CMIP5 AMIP experiment with those in T-AMIP2 and use a case-study approach to gain a greater understanding of possible causes.

In the next section we describe the T-AMIP2 experiment and diagnostics collected. The models and observational data used in the subsequent analysis are presented in Section 3. In Section 4 the meteorological situations in which the biases are largest in the AMIP and T-AMIP2 experiments are discussed. Results from a T-AMIP2 case study which has typical meteorological conditions in which the bias is present are shown in Section 5, whilst conclusions are given in Section 6.

2 Transpose-AMIP II

a Experimental design

The latest information on T-AMIP2 is available from http://www.transpose-amip.info. Here we document the key details of the T-AMIP2 experimental design.

T-AMIP2 comprises 64 hindcasts using a center's Atmosphere GCM (AGCM), each hindcast being five days in length. The hindcasts are split into four sets. The full list of hindcast start times are given in Table 1. The hindcast start times in each set are at 30

hour intervals to ensure sampling throughout the diurnal cycle for each grid-point for a given lead time. This is particularly important for some diagnostics which are only available at sunlit points. The 2008/9 period was chosen to tie in with the Year of Tropical Convection (YOTC; Waliser *et al*, 2012) during which ECMWF analyses have been made generally available to the research community, and various field campaigns and other modeling studies are being undertaken. The four sets of hindcasts were chosen to evenly sample the annual cycle whilst also providing several hindcasts within one or more of the Intensive Observation Periods (IOPs) for VOCALS (VAMOS (Variability of the American Monsoon Systems) Ocean-Cloud-Atmosphere-Land Study), AMY (Asian Monsoon Years) and T-PARC (THORPEX (THe Observing-system Research and Predictability EXperiment) Pacific Asian Regional Campaign).

Model state variables are initialized from ECMWF YOTC analyses. Centers are advised to interpolate the analysis onto their model grid following the ECMWF Integrated Forecast System documentation: http://www.ecmwf.int/research/ifsdocs_old/TECHNICAL/. Atmospheric composition, solar forcing and land use are as 2008/9 of the CMIP5 AMIP experiment. It is recommended that sea surface temperatures from the ECMWF YOTC analyses are persisted in the hindcasts.

The longer timescale over which processes in land surface models operate, compared with the atmosphere, means that a straight transplant of initial conditions from an analysis produced by a different land surface model may not always be appropriate (e.g. Boyle *et al*, 2005). However, feedback from modeling centers in the planning stages of T-AMIP2 indicated that being overly prescriptive on the land surface initialization may deter centers from participating. Whilst for some studies, the choice of land surface initialization will be important, for others (such as the study presented in this paper), it is less so. It

was therefore decided to allow a choice of land surface initialization from the following methods, but requesting the participants clearly indicate which methodology they have used:

- Initialize from fields produced by a land surface assimilation system (e.g. ECMWF or Global Land Data Assimilation Systems (GLDAS)).
- Initialize using a suitable climatology: e.g. from Global Soil Wetness Project 2 (GSWP2) or derived from the model's AMIP simulation.
- Initialized with a nudging method as described by Boyle *et al* (2005).

Aerosol concentrations are either initialized using a climatology calculated from the model's AMIP simulation, or initialized using the nudging method of Boyle *et al* (2005). Non-state variable prognostics which spin-up quickly (such as cloud fraction for models with a prognostic scheme) can either be initialized from zero, or initialized using the nudging method of Boyle *et al* (2005).

In order to use the T-AMIP2 experiment to comment on the processes operating in the climate change simulations of CMIP5, the AGCMs submitted to T-AMIP2 are requested to be identical, both in terms of the model science and resolution, to those used for the CMIP5 AMIP experiment. Whilst maximum benefit from the project is expected for climate models, T-AMIP2 is also open to NWP centers to participate. WGNE encouraged NWP centers to run and submit an AMIP simulation for CMIP5 and any centers undertaking this experiment are asked to also submit to T-AMIP2.

In order to test the sensitivity of the results to the choice of analysis, centers are given the option to also submit a second set of T-AMIP2 hindcasts using an alternative initialization. For this optional set of hindcasts, centers with their own assimilation system

will use their own analyses whilst centers without their own assimilation system are asked to initialize from the Global Modeling and Assimilation Office Modern-Era Retrospective analysis for Research and Applications (GMAO/MERRA).

b Requested diagnostics and data access

A key aim of T-AMIP2 is that it should allow detailed diagnostic analysis of the processes operating in the model. As such a comprehensive set of diagnostics are requested at high temporal resolution. All of the requested diagnostics appear within the CMIP5 diagnostic lists, so no new diagnostics need to be especially produced for T-AMIP2.

The full details of the requested diagnostics can be found under the 'Data requirements' section of http://www.transpose-amip.info, but they include:

- All single level and multi-level (both model level and standard pressure level) fields which are usually collected as monthly means in a AMIP experiment are requested as 3-hour means through the T-AMIP2 hindcasts.

- All fields normally used for standard NWP verification (e.g. in calculating Commission for Basic Systems (CBS) scores).

- 3-hour mean temperature and humidity tendencies on model levels from various sections of the model physics/dynamics.

- Output from the CFMIP (Cloud Feedback Model Intercomparison Project) Observation Simulator Package (COSP; Bodas-Salcedo *et al*, 2011) which produces model diagnostics which emulate several satellite products including ISCCP (International Satellite Cloud Climatology Project), CloudSat and CALIPSO (Cloud Aerosol Lidar and Infrared Pathfinder Satellite Observations). CloudSat and CALIPSO are

part of the NASA (National Aeronautics and Space Administration) A-train, and T-AMIP2 requests curtain data from the models which matches the satellite orbit.

- At 119 sites identified by the CFMIP project, many of the above diagnostics, including model tendencies, are requested at 30 minute intervals through the hindcast, which for many climate models is close (or equal) to the model timestep. This allows detailed examination of the model evolution through the hindcast. The sites are listed at http://www.cfmip.net and include ARM and CloudNet (Illingworth *et al*, 2007) sites, as well as points along transects including the GCSS (GEWEX (Global Energy and Water Cycle Experiment) Cloud Systems Study) Pacific cross section (GPCI), the VOCALS cross section and the AMMA (African Monsoon Multidisciplinary Analyses) cross section.

The format of, and access to, data from T-AMIP2 follows CMIP5 with all submitted data being CF (Climate-Forecast)–compliant netCDF (network Common Data Form) and conforming to the standards of CMOR (Climate Model Output Rewriter; Taylor *et al*, 2012). T-AMIP2 data are freely available for research purposes and can be downloaded from the Earth System Grid Federation (ESGF) (e.g. http://pcmdi3.llnl.gov/esgcet), identical to CMIP5 except for selecting the project to be 'TAMIP'.

3 Models and observational data

At the time of submission (5th July 2012), output from five models had been submitted to T-AMIP2 and those models are used in this study:

- NCAR/CCSM4 (National Center for Atmospheric Research/Community Climate System Model version 4) (Gent *et al*, 2011; Neale *et al*, 2010) is a grid-point model

with a horizontal resolution of N144 (approx. $90km$ in the mid-latitudes) and 26 vertical levels. (The 'N' notation denotes half the number of longitudinal grid-points. Since a wave cannot be represented with less than 2 grid-points, this allows approximate comparison with 'T' notation for spectral models.)

- CNRM-CERFACS/CNRM-CM5 (Centre National de Recherches Météorologiques - Centre Européen de Recherche et de Formation Avancée en Calcul Scientifique/CNRM - Coupled Model version 5) (Voldoire *et al*, 2012) is a spectral model with a horizontal resolution of TL127 (approx. $160km$ uniformly over the globe) and 31 vertical levels.

- MOHC/HadGEM2-A (Met Office Hadley Centre/Hadley Centre Global Environmental Model version 2 - Atmosphere) (Martin *et al*, 2011) is a grid-point model with a horizontal resolution of N96 (approx. $135km$ in the mid-latitudes) and 38 vertical levels.

- IPSL/IPSL-CM5A-LR (Institut Pierre Simon Laplace - Climate Model version 5A - Low Resolution) (http://icmc.ipsl.fr) is a grid-point model with a horizontal resolution of N48 (approx. $270km$ in the mid-latitudes) and 39 vertical levels.

- MIROC/MIROC5 (Model for Interdisciplinary Research On Climate version 5) (Watanabe *et al*, 2010) is a spectral model with a horizontal resolution of T85 (approx. $150km$ in the mid-latitudes) and 40 vertical levels.

A range of observational data is used in this study. Mean top of atmosphere (TOA) fluxes from AMIP simulations are compared with CERES-EBAF (Clouds and the Earth's Radiant Energy System - Energy Balanced And Filled) data (Loeb *et al*, 2009). Daily TOA fluxes are evaluated against CERES-Flashflux (Wielicki and Coauthors, 1996),

ISCCP-FD (Zhang *et al*, 2004) and ERBE (Earth Radiation Budget Experiment; Barkstrom and Smith, 1986) datasets.

Cloud comparisons are made against the ISCCP D1 dataset (Rossow and Schiffer, 1999) which comprises 3-hourly histograms of cloud fraction on a 2.5^o grid in 7 cloud top pressure and 6 optical depth bins. The lidar backscatter from clouds along a section of the A-train orbit are are presented from CALIPSO (Winker *et al*, 2010). This uses a nadir-pointing instrument with a beam diameter of $70m$ at the Earths surface and produces footprints every $333m$ in the along-track direction. We use the GCM Oriented CALIPSO Cloud Product (Chepfer *et al*, 2010) and comparison with the model is through height-scattering ratio histograms of cloud amount.

4 Nature of the Southern Ocean cloud biases

The climatological bias in TOA reflected shortwave (RSW) and outgoing longwave (OLR) over the Southern Ocean region is obtained from the CMIP5 AMIP experiment. Here, we focus on those models which have submitted both their AMIP and T-AMIP2 experiments (Figure 1). During most of the year, the bias in RSW is considerably larger than that in OLR. The models show the largest bias in RSW during the Austral summer when insolation is at a maximum. For three of the models (CNRM-CM5, HadGEM2-A and MIROC5), there is a negative bias in RSW which reaches a peak of -30 – -45Wm^{-2} between 60^oS and 65^oS. This is a common bias amongst many of the CMIP5 AMIP simulations and is qualitatively consistent with the CMIP3 coupled model results presented in Trenberth and Fasullo (2010). CCSM4 also shows a negative bias in RSW, but covering a broader range of latitudes with a peak much further north (45^oS). IPSL-CM5A-LR has a positive bias in RSW over the region, peaking at 50^oS and, across the region as a whole,

has the largest positive RSW bias. Differences in the clear-sky RSW (not shown) between these models is small, suggesting clouds are largely responsible for these biases. In this paper we focus on the common negative RSW bias between 60^oS and 65^oS.

The meteorology of the Southern Ocean is characterized by relatively fast moving synoptic depressions with their associated frontal systems, interspersed with transient ridges. It is difficult to establish from climatological means in which of these synoptic conditions the RSW bias is most prevalent. We therefore follow the cyclone compositing methodology of Field and Wood (2007), recently applied to Southern Ocean clouds by Bodas-Salcedo *et al* (2012), to establish the location of the maximum bias in an average cyclone. Cyclone centers are identified from five summers (DJF) of daily mean sea-level pressure (MSLP) fields from the AMIP simulation. A box covering 60^o longitude and 30^o latitude centered on the cyclone is identified with the model RSW field over the region of the box being saved. The individual cyclone boxes are then averaged. Since these boxes cover a reasonably large area, the vast majority of grid-points are included at least once in the cyclone composite. We also conduct the cyclone compositing on MSLP data from ECMWF re-analyses and use this in conjunction with the RSW fields from ISCCP-FD and ERBE to form an observed RSW cyclone composite. The difference between the model and observed mean RSW fields provides the mean bias around the composite cyclone (Figure 2). The position of fronts on any one cyclone on an individual day will vary, but the schematic on Figure 2 shows the typical location.

For CNRM-CM5, HadGEM2-A and MIROC5, the negative bias in the AMIP RSW exists around much of the composite cyclone (except the cyclone center). It is a maximum within the south-westerly flow on the cold air side of the cyclone, which could also be on the leading side of a following transient ridge. The bias is smaller in the region typically

occupied by fronts and near the cyclone center suggesting that these cloud systems are not generally responsible for the negative RSW bias. This picture of the bias being a maximum on the cold air side of the cyclone away from fronts is consistent with the findings of Bodas-Salcedo *et al* (2012) when examining the Met Office model. The cyclone composite RSW bias pattern for IPSL-CM5A-LR is structurally similar but systematically more positive with the positive RSW bias in this model being mainly associated with frontal systems with a near-zero bias in the cold-air region.

The right-hand column of Figure 2 shows the RSW bias from a similar cyclone compositing of the second day from each member of the Jan/Feb 2009 set of hindcasts from T-AMIP2. Although the plots are more noisy due to only 16 days being considered rather than circa 450 from the AMIP simulation, the same picture of a maximum in the negative bias in RSW on the cold-air side of the cyclone emerges, indicating that the bias develops early in the model simulation and can be investigated using the T-AMIP2 experiments.

The T-AMIP2 cyclone composite for CCSM4 shows one of the largest negative RSW biases amongst the models (the required daily diagnostics were not available from this model for the AMIP simulations). Given that the location of the cyclones will be well constrained in the T-AMIP2 experiments, this is perhaps surprising since the mean RSW bias for CCSM4 is smaller than for the other GCMs in the storm track region of Figure 1c. This can be understood from Figure 3 which shows the mean evolution in RSW during the Jan/Feb 2009 hindcasts. Since the models are starting from ECMWF analyses they will typically need to 'spin-up' cloud. This happens relatively quickly and most models achieve their final mean RSW within a few hours, and certainly within the first day. The exception is CCSM4 which takes around 3 days for the RSW to reach equilibrium and on the second day (when the analysis for Figure 2 is undertaken) has an RSW bias between

60^oS and 65^oS which is comparable with most of the other models. Given the similarity with the other models in Figure 2, we suggest that CCSM4 develops the same negative bias early on in the hindcasts, but there are then additional processes which further affect the cloud (possibly a northward shift of the storm track), which occur over several days and result in the different climatological RSW bias seen in Figure 1c. A longer timescale for the development of climatological cloud biases in the NCAR model was also noted by Zhang *et al* (2010) and Medeiros *et al* (2012), in contrast to the rapid development of the cloud biases in the Met Office model (Williams and Brooks, 2008).

To further establish the meteorological conditions under which the negative RSW bias is present, we composite each model grid-point from the second day of each of the Jan/Feb 2009 hindcasts according to the model $10m$ wind speed and a measure of lower tropospheric stability (LTS). Here we follow Williams *et al* (2006) by using the difference in saturated equivalent potential temperature (θ_{es}) between $700hPa$ and the surface. Figure 4 shows the mean RSW bias in each bin weighted by the total area of all the grid-points included in that bin, such that averaging across the histograms will give the mean RSW bias from the models for the Southern Ocean region.

There is a clear dependence of the RSW bias on the LTS with the strong negative bias in most models being present in stable conditions, and the positive bias in IPSL-CM5A-LR being a maximum when the LTS is around zero. Although more stable lower tropospheres may generally be considered to be associated with calm conditions, it can be seen that the negative RSW bias occurs across a range of wind speeds and, in the more stable situations where the LTS is greater than $10K$ in CNRM-CM5 and MIROC5, the maximum bias is associated with relatively strong daily-mean $10m$ wind speeds of around $10ms^{-1}$.

5 T-AMIP2 case study analysis

a Cloud evaluation

We have searched through the second day of the Jan/Feb T-AMIP2 hindcasts to identify typical cases matching the meteorological conditions under which the RSW bias is a maximum. Several have been identified (not shown), of which a region of the Atlantic sector of the Southern Ocean on 17th Jan 2009 is typical and will be analyzed in detail here since there was also a pass of the A-train which coincided with the bias being present. Cases similar to the one presented here occurred frequently in the hindcasts with the RSW biases being very similar.

Figure 5 shows the Met Office synoptic analysis for 12Z on 17th Jan 2009. A cold front has just passed the prime meridian associated with a depression centered near 24^oW 74^oS. Behind is a region of cold advection in a strong south-westerly flow and forming the leading side of a transient ridge.

The MSLP from each model for the T-AMIP2 hindcast initialized at 06Z Jan 16th and verifying at 12Z Jan 17th 2009 is generally in good agreement with the synoptic analysis (Figure 6). This indicates that the large-scale dynamics are generally well constrained for this analysis as would be expected 30 hours into a hindcast. There is a large negative RSW bias for the second day (Jan 17th) of this hindcast in all of the T-AMIP2 models in the region behind the cold front (marked with a dotted box) which will be investigated here. Despite its overall positive RSW bias, IPSL-CM5A-LR is consistent with the other GCMs in having a negative bias for the case study region, although it is smaller in magnitude than for the other models. The region marked with the box is reasonably typical of the meteorological conditions associated with a high RSW bias identified in Figure 2 (post cold

front in a region of south-westerly flow) and Figure 4 (stable conditions with a reasonably strong surface flow). The location of this region, centered at 61^oS, is also consistent with the latitude of the largest negative RSW bias in Figures 1c & 3b.

In addition to the core set of hindcasts initialized from ECMWF analyses, the T-AMIP2 experimental design has an option to also submit a second set of hindcasts initialized from alternative analyses. HadGEM2-A has also been submitted with a set of hindcasts initialized from Met Office analyses. Whilst there are some differences in the RSW bias, for the region being investigated here there is very little difference between the hindcasts, indicating that the choice of analysis is not a factor in the development of this bias (Figure 6).

The mean cloud fraction histogram from ISCCP for 12Z on Jan 17th over the dotted box in Figure 6 is shown in Figure 7, together with the comparable COSP output from the models 30 hours into the hindcast and verifying at the same time as the observations. There is 100% cloud cover over the region in the observations, with a large proportion having tops between $560hPa$ and $680hPa$ and optical depths between 23.0 and 60.0. The models have slightly lower cloud fractions than observed (much lower in the case of CCSM4), and all have smaller optical depths (typically between 3.6 and 9.4) with the low-cloud tops being mostly at a lower altitude.

An alternative evaluation of the cloud may be obtained from instruments on the A-train which passed over the region along the dashed line in Figure 6 at around 14:20UTC on Jan 17th. The cloud radar on CloudSat is generally not able to resolve cloud below $1km$ in altitude due to ground clutter. There was also no signal from the low cloud above $1km$ present in this case, suggesting that it was non-precipitating. Instead we compare the model with data from the CALIPSO lidar. Figure 8a shows the lidar backscatter ratio

along the transect, cutting the cold front at either end of the cross section. Processing of the lidar data has a very tight solar contamination threshold, and because this is a daytime pass there are many 'missing data' columns in the curtain data. However, a layer of continuous low cloud can still be clearly seen through the center of the transect with the cloud top altitude decreasing from around $3km$ at the southern end to $1.5km$ at the northern end.

The central section of this transect has been processed into an altitude – scattering ratio histogram, (Figure 8b). The lidar is sensitive to optically very thin cloud and the histogram shows a small amount of very thin cloud is present on occasions through much of the troposphere. However the majority of the cloud in the central region of the transect has cloud tops between $1.5km$ and $3km$ with scattering ratios in excess of 80 indicating that the cloud is highly reflective. This is consistent with the high optical depths seen in the ISCCP observations (Figure 7) and the agreement between these two independent, and quite different, observations provides considerable confidence in the cloud properties present at the time. Again, consistent with the COSP diagnostics from the models used for the ISCCP comparison, the model histograms produced by COSP for CNRM-CM5 and HadGEM2-A show the cloud top to be too low in the models and too optically thin, indicated by the lower scattering ratio than observed (Figure 8c&d).

b Boundary layer structure

Given that the the models appear to generally share a bias of the boundary layer cloud being too low and insufficiently reflective, we now examine the boundary layer structure at a point 2^oW 61^oS and marked with the asterisk in Figure 6.

The profiles of temperature and humidity from the ECMWF analysis at 12Z 17th Jan

2009 is plotted in Figure 9. The profile indicates a near saturated layer between $925hPa$ and $870hPa$. Above this is a strong temperature inversion with its top at $815hPa$ and a much dryer free troposphere above this. The profiles from the models verifying at the same time, 30 hours into the hindcast, generally capture the presence of an inversion with a dryer free troposphere, although the detail is often in error. In particular, all of the T-AMIP2 models have an inversion which is too low. Several of the models have a near saturated layer below the inversion which is less than half the physical thickness observed. In contrast, CNRM-CM5 (and to a lesser extent CCSM4) appears to have a warmer boundary layer and so be further from saturation, possibly accounting for the lower scattering ratio in Figure 8 and higher RSW bias in Figure 6 compared with the other models. IPSL-CM5A-LR, although having a cooler boundary layer than the EC analysis, has the thickest layer close to saturation of the T-AMIP2 models and has the highest inversion base of these models. This may well be associated with the smaller RSW bias for IPSL-CM5A-LR in Figure 6.

These common biases in the boundary layer structure are consistent with the cloud biases and indicates that the inversion being too low is related to the cloud being lower than observed. It is also possible that the saturated layer being physically thinner than observed may, at least in part, be associated with the cloud optical depths and scattering ratios being too low.

To investigate the development of these errors in the boundary layer structure we make use of the temperature tendency diagnostics requested by T-AMIP2. We now examine the early stages of the hindcast initialized at 12Z on Jan 17th 2009 (the verification time of profiles shown in Figure 9). Since the models are starting from an alien analysis, there is often a significant adjustment of schemes in the first timestep whilst the non-

initialized prognostic variables (e.g. cloud) spin up. This can affect the first 3-hour mean tendency, but subsequent tendencies are found to be more consistent and show some common behaviors between the models. Figure 10 shows the mean temperature tendencies between 3 and 6 hours into the hindcast. The total temperature tendency (solid black) indicates that the models all have a warming tendency at altitudes above $890hPa$. The level that this warming extends down to varies between the models from $890hPa$ in IPSL-CM5A-LR to the surface in CCSM4. The other models have a slight cooling tendency near the surface. During this period the environment in the real world will be evolving and so we take the difference between successive ECMWF analyses to estimate the 'observed' tendency for the point (dashed black). This also shows a warming peaking at around $870hPa$ but this changes to a cooling at altitudes below $890hPa$. Hence we have a picture that during the afternoon of Jan 17th, the inversion at this point is becoming stronger, but in CCSM4, HadGEM2-A and MIROC5 the warming is extending too far down. IPSL-CM5A-LR is closer to the 'observed' tendency, consistent with the higher inversion base for this model in Figure 9.

T-AMIP2 requests tendency diagnostics from various sections of the model science, enabling the evolution of the total tendency to be examined. These are shown colored in Figure 10 and the gray line shows their sum. This sum may not always be identical to the total tendency due to the effect of other, model specific, aspects of the model science on the temperature, but should usually be close. It can be seen that much of the anomalous warming in MIROC and HadGEM2-A below $890hPa$ is coming through the advection. It should be noted though that the error in the total tendency may be due to an insufficient compensating cooling effect from other processes. Indeed, this is suggested in IPSL-CM5A-LR where there is a similar warm advection around $900hPa$, but it is being

balanced by radiative cooling. It is likely that this radiative cooling is coming from the thicker cloud sheet in this model, hence there appears to be an important and sensitive feedback between the cloud and the boundary layer structure in the first few hours of the model simulation. If the cloud layer is not maintained sufficiently strongly then the boundary layer structure will evolve, thinning the cloud further.

Given this sensitivity, accurate representation of the details of the boundary layer structure are likely to be important, but the vertical resolution of these models is too coarse to resolve the details of the inversion. The position of the model levels has been marked on the right side of each panel in Figure 10. IPSL-CM5A-LR appears to maintain the cloud layer and inversion height with relatively coarse resolution, however even this model has errors in the temperature structure, being too cold in the boundary layer (Figure 9). For HadGEM2-A in particular, it can be seen that the lack of vertical resolution may be a factor in the lower inversion height since the temperature tendencies at the model level points are close to those observed, but the spacing of the levels means that the gradient between the warming and cooling is not strong enough (Figure 10). We have re-run the hindcast initialized at 6Z on Jan 16th with a more recent configuration of the Met Office Unified Model (MetUM) using the GA4 (Global Atmosphere 4) science (Walters et al., in prep.). Many aspects of this model have changed since HadGEM2-A including a new cloud scheme and significant developments to the boundary layer scheme. However one of the largest impacts on the simulation of cloud globally was the introduction of higher vertical resolution (from 38 to 85 levels). In particular, MetUM GA4 has almost twice the number of levels below $700hPa$ as HadGEM2-A. The hindcast has been re-run following the T-AMIP2 protocol, initialized from ECMWF analyses and run at the same horizontal resolution as HadGEM2-A.

Temperature and humidity profiles for MetUM GA4 30 hours into the hindcast have been added to Figure 9. Whilst there are still differences with respect to the ECMWF analyses, it can be seen that the model captures the boundary layer structure - the inversion and near-saturated layer below - much better than HadGEM2-A. It is very likely that improved representation of the inversion in this case has been enabled through the higher vertical resolution. Consistent with this improved boundary layer structure is that the RSW bias has been almost halved for the region being studied here (cf. Figure 11 with HadGEM2-A in Figure 6).

6 Conclusions

In this paper we have presented the T-AMIP2 experiment: a coordinated model intercomparison project, running alongside CMIP5, in which climate models are run in 'weather forecast mode', initialized from a common analysis. The aim of the project is to permit detailed diagnostic investigations into the processes operating in models being used for climate change projections in CMIP5. In order to fulfill this aim, T-AMIP2 requests a comprehensive set of diagnostics including satellite simulator and tendency diagnostics at high temporal resolution, plus near-timestep diagnostics at a selection of sites. The set of hindcasts have been chosen to tie in with IOPs within a number of field campaigns, and is all set within the intensively studied YOTC period.

The use of some of these diagnostics for detailed investigations has been illustrated in an analysis of TOA flux biases over the Southern Ocean. This is an issue for many GCMs and is believed to affect their coupled atmosphere-ocean performance and climate change response. Most of the models submitted to T-AMIP2 share a bias of too little RSW over the Southern Ocean, although details vary and IPSL-CM5A-LR is notably different in

having a positive bias. The negative RSW bias has been shown to be present primarily on the cold-air side of cyclones/leading side of transient ridges, away from frontal regions. The bias is present when the lower troposphere is more stable, but can be associated with a range of low level wind speeds.

A case study has been presented which is typical of the conditions under which an RSW bias is present in all of the models. Generally the biases are more similar between the models at these short ranges. Variations in the climatological bias then develop due to differences in the magnitude of positive and negative biases in different synoptic conditions and, in some models, due to longer timescale feedbacks. Investigation of the T-AMIP2 hindcasts for the case study has revealed the bias to develop quickly (within the first day) and be primarily due to the cloud being too optically thin, with an additional contribution from the cloud fraction being too low in some models. It appears that the overly thin stratocumulus is associated with the inversion being quickly lowered at the beginning of the hindcast and a physically thinner cloud layer being produced.

This type of boundary layer appears to be problematic for all of the models investigated and examination of model tendencies over the first few hours to the hindcast suggests the GCMs can be sensitive to a positive feedback process between the removal of cloud and evolution of the boundary layer structure. Details of the exact cause of the errors in the boundary layer structure and cloud are likely to vary from model to model and could well have more than one source. However identifying the conditions under which models develop the bias should assist model developers in each center in investigating further and focus improvements to the science in their model. In at least one model, the lack of sufficient vertical resolution to properly represent the boundary layer temperature and humidity structure appears to be a factor. In a more recent configuration of this

model, with almost double the vertical resolution in the boundary layer, the RSW bias is significantly reduced.

This is the first T-AMIP2 study, however there are a very large number of other studies which could potentially be carried out with T-AMIP2 data. The data are freely available for use by the research community and we encourage scientific researchers to conduct their own investigations. Several modeling centers have indicated that they intend to submit T-AMIP2 data over the coming year and we are confident that more will follow. We believe that a set of detailed diagnostic investigations using these data will lead to a greater understanding of which processes we have confidence in and which require a more focused effort to improve in the current generation of climate models.

Acknowledgments

This work was supported by the Joint DECC/Defra Met Office Hadley Centre Climate Programme (GA01101).

The contribution of B. Medeiros to this work was supported by the Office of Science (BER), U.S. Department of Energy, Cooperative Agreement DE-FC02-97ER62402. NCAR is sponsored by the National Science Foundation.

The contribution of S.A. Klein to this work was supported by the Regional and Global Climate Modeling and Atmospheric System Research Programs of the Office of Science at the U.S. Department of Energy and was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.

We acknowledge WGNE and WGCM who are responsible for T-AMIP2 and CMIP5, and we thank the climate modeling groups listed in Section 3 for producing and making

516 available their model output. For CMIP5 and T-AMIP2 the U.S. Department of En-
517 ergy's Program for Climate Model Diagnosis and Intercomparison provides coordinating
518 support and led development of software infrastructure in partnership with the Global
519 Organization for Earth System Science Portals.

520 CERES, ERBE and ISCCP data were obtained from the NASA Langley Research
521 Center Atmospheric Sciences Data Center. CALIPSO-GOCCP data were obtained from
522 the IPSL ClimServ data centre (http://climserv.ipsl.polytechnique.fr/cfmip-obs.html).

References

Barkstrom, B. R., and G. L. Smith, 1986: The earth radiation budget experiment: Science and implementation. *Rev. Geophys.*, **24**, 379–390.

Bodas-Salcedo, A., M. J. Webb, S. Bony, H. Chepfer, J. L. Dufresne, S. Klein, Y. Zhang, R. Marchand, J. M. Haynes, R. Pincus, and V. O. John, 2011: COSP: satellite simulation software for model assessment. *Bull. Am. Meteorol. Soc.*, **92**(8), 1023–1043. doi:10.1175/2011BAMS2856.1.

——, K. D. Williams, P. R. Field, and A. P. Lock, 2012: The surface downwelling solar radiation surplus over the Southern Ocean in the Met Office model: the role of midlatitude cyclone clouds. *J. Climate*. In press. doi:10.1175/JCLI-D-11-00702.1.

Boyle, J., and S. A. Klein, 2010: Impact of horizontal resolution on climate model forecasts of tropical precipitation and diabatic heating for the TWPICE period. *J. Geophys. Res.*, **115**(D23113). doi:10.1029/2010JD014262.

Boyle, J. S., D. Williamson, R. Cederwall, M. Fiorino, J. Hnilo, J. Olson, T. Phillips, G. Potter, and S. Xie, 2005: Diagnosis of Community Atmospheric Model 2 (CAM2) in numerical weather forecast configuration at Atmospheric Radiation Measurement sites. *J. Geophys. Res.*, **110**(D15S15). doi:10.1029/2004JD005042.

Brown, A., S. Milton, M. Cullen, B. Golding, J. Mitchell, and A. Shelly, 2012: Unified modelling and prediction of weather and climate: a 25 year journey. *Bull. Am. Meteorol. Soc.* Submitted.

Chepfer, H., S. Bony, D. Winker, G. Cesana, J.-L. Dufresne, P. Minnis, C. J. Stubenrauch, and S. Zeng, 2010: The GCM Oriented Calipso Cloud Product (CALIPSO-GOCCP). *J. Geophys. Res.*, **115**, D00H16. doi:10.1029/2009JD012251.

Field, P. R., and R. Wood, 2007: Precipitation and cloud structure in midlatitude cyclones. *J. Climate*, **20**(2), 233–254. doi:10.1175/JCLI3998.1.

Gates, W. L., and 15 coauthors, 1999: An overview of the results of the Atmospheric Model Intercomparison Project (AMIP I). *Bull. Am. Meteorol. Soc.*, **80**(1), 29–55.

533 Gent, P. R., G. Danabasoglu, L. J. Donner, M. M. Holland, E. C. Hunke, S. R. Jayne, D. M. Lawrence, R. B. Neale, P. J. Rasch, M. Vertenstein, P. H. Worley, Z.-L. Yang, and M. Zhang, 2011: The community climate system model version 4. *J. Climate*, **24**, 4973–4991. doi:10.1175/2011JCLI4083.1.

534 Haynes, J. M., C. Jakob, W. B. Rossow, G. Tselioudis, and J. Brown, 2011: Major characteristics of Southern Ocean cloud regimes and their effects on the energy budget. *J. Climate*, **24**, 5061–5080. doi:10.1175/2011JCLI4052.1.

535 Illingworth, A. J., R. J. Hogan, E. J. O'Connor, D. Bouniol, M. E. Brooks, J. Delano, D. P. Donovan, J. D. Eastment, N. Gaussiat, J. W. F. Goddard, M. Haeffelin, H. K. Baltink, O. A. Krasnov, J. Pelon, J. M. Piriou, A. Protat, H. W. J. Russchenberg, A. Seifert, A. M. Tompkins, G. J. van Zadelhoff, F. Vinit, U. Willn, D. R. Wilson, and C. L. Wrench, 2007: CLOUDNET: continuous evaluation of cloud profiles in seven operational models using ground-based observations. *Bull. Am. Meteorol. Soc.*, **88**, 883–898. doi:10.1175/BAMS-88-6-883.

536 Loeb, N. G., B. A. Wielicki, D. R. Doelling, S. Kato, T. Wong, G. L. Smith, D. F. Keyes, and N. Manalo-Smith, 2009: Toward optimal closure of the Earth's top-of-atmosphere radiation budget. *J. Climate*, **22**(3), 748–766. doi:10.1175/2008JCLI2637.1.

537 Martin, G. M., S. F. Milton, C. A. Senior, M. E. Brooks, S. Ineson, T. Reichler, and J. Kim, 2010: Analysis and reduction of systematic errors through a seamless approach to modelling weather and climate. *J. Climate*, **23**, 5933–5957. doi:10.1175/2010JCLI3541.1.

538 ——, N. Bellouin, W. J. Collins, I. D. Culverwell, P. R. Halloran, S. C. Hardiman, T. J. Hinton, and C. D. Jones, 2011: The HadGEM2 family of Met Office Unified Model climate configurations. *Geosci. Model Devel.*, **4**, 723–757. doi:10.5194/gmd-4-723-2011.

539 Medeiros, B., D. L. Williamson, C. Hannay, and J. G. Olson, 2012: Southeast Pacific stratocumulus in the Community Atmosphere Model. *J. Climate*. In press. doi:10.1175/JCLI-D-11-00503.1.

540 Neale, R. B., J. H. Richter, A. J. Conley, S. Park, P. H. Lauritzen, A. Gettelman, D. L. Williamson, P. J. Rasch, S. J. Vavrus, M. A. Taylor, W. D. Collins, M. Zhang, and S.-J. Lin, 2010: Description of the NCAR Community Atmosphere Model (CAM 4.0). Technical Report NCAR/TN-485+STR, NCAR Technical Note. Available from http://www.cesm.ucar.edu/models/ccsm4.0/cam/.

541 Phillips, T. J., G. L. Potter, D. L. Williamson, R. T. Cederwall, J. S. Boyle, M. Fiorino, J. J. Hnilo, J. G. Olson, S. Xie, and J. J. Yio, 2004: Evaluating parameterizations in General Circulation Models: Climate simulation meets weather prediction. *Bull. Am. Meteorol. Soc.*, **85**, 1903–1915.

542 Rodwell, M. J., and T. N. Palmer, 2007: Using numerical weather prediction to assess climate models. *Q. J. R. Meteorol. Soc.*, **133**, 129–146. doi:10.1002/qj.23.

543 Rossow, W. B., and R. A. Schiffer, 1999: Advances in understanding clouds from ISCCP. *Bull. Am. Meteorol. Soc.*, **80**, 2261–2287.

544 Senior, C. A., A. Arribas, A. R. Brown, M. J. P. Cullen, T. C. Johns, G. M. Martin, S. F. Milton, S. Webster, and K. D. Williams, 2010: *Synergies between numerical weather prediction and general circulation climate models.*, chapter In: The Development of Atmospheric General Circulation Models: Complexity, Synthesis, and Computation (Donner L., Somerville R., Schubert W., eds). Cambridge University Press.

545 Taylor, K. E., R. J. Stouffer, and G. A. Meehl, 2012: An overview of CMIP5 and the experiment design. *Bull. Am. Meteorol. Soc.*

546 Trenberth, K. E., and J. T. Fasullo, 2010: Simulation of present-day and twenty-first-century energy budgets of the southern oceans. *J. Climate*, **23**(2), 440–454. doi:10.1175/2009JCLI3152.1.

547 Voldoire, A., E. Sanchez-Gomez, D. Salas-y Melia, B. Decharme, C. Cassou, S. Senesi, S. Valcke, I. Beau, A. Alias, M. Chevallier, M. Deque, J. Deshayes, H. Douville, E. Fernandez, G. Madec, E. Maisonnave, M.-P. Moine, S. Planton, D. Saint-Martin, S. Szopa, S. Tyteca, R. Alkama, S. Belamari, A. Braun, L. Coquart, and F. Chauvin, 2012: The CNRM-CM5.1 global climate model: description and basic evaluation. *Clim. Dyn.* doi:10.1007/s00382-011-1259-y.

548 Waliser, D. E., M. Moncrieff, D. Burrridge, A. Fink, D. Gochis, N. Goswami, B. Guan, P. Harr, J. Heming, H.-H. Hsu, C. Jakob, M. Janiga, R. Johnson, S. Jones, P. Knippertz, J. Marengo, H. Nguyen, M. Pope, Y. Serra, C. Thorncroft, M. Wheeler, R. Wood, and S. Yuter, 2012: The "Year" of Tropical Convection (May 2008 to April 2010): Climate variability and weather highlights. *Bull. Am. Meteorol. Soc.* In press. doi:10.1175/2011BAMS3095.1.

549 Watanabe, M., T. Suzuki, R. Oishi, Y. Komuro, S. Watanabe, S. Emori, T. Takemura, M. Chikira, T. Ogura, M. Sekiguchi, K. Takata, D. Yamazaki, T. Yokohata, T. Nozawa, H. Hasumi, H. Tatebe, and M. Kimoto, 2010: Improved climate simulation by MIROC5: Mean states, variability, and climate sensitivity. *J. Climate*, **23**, 6312–6335. doi:10.1175/2010JCLI3679.1.

550 Wielicki, B., and Coauthors, 1996: Clouds and the Earth's Radiant Energy System (CERES): an earth observing system experiment. *Bull. Am. Meteorol. Soc.*, **77**(5), 853–868.

551 Williams, K. D., and M. E. Brooks, 2008: Initial tendencies of cloud regimes in the Met Office Unified Model. *J. Climate*, **21**(4), 833–840. doi:10.1175/2007JCLI1900.1.

552 ——, M. A. Ringer, C. A. Senior, M. J. Webb, B. J. McAvaney, N. Andronova, S. Bony, J.-L. Dufresne, S. Emori, R. Gudgel, T. Knutson, B. Li, K. Lo, I. Musat, J. Wegner, A. Slingo, and J. F. B. Mitchell, 2006: Evaluation of a component of the cloud response to climate change in an intercomparison of climate models. *Clim. Dyn.*, **26**, 145–165. doi:10.1007/s00382-005-0067-7.

553 Williamson, D. L., J. Boyle, R. Cederwall, M. Fiorino, J. Hnilo, J. Olson, T. Phillips, G. Potter, and S. C. Xie, 2005: Moisture and temperature balances at the atmospheric radiation measurement southern great plains site in forecasts with the Community Atmosphere Model (CAM2). *J. Geophys. Res.*, **110**(D15S16), 17.

554 Winker, D. M., J. Pelon, J. A. Coakley Jr, S. A. Ackerman, R. J. Charlson, P. R. Colarco, P. Flamant, Q. Fu, R. M. Hoff, C. Kittaka, T. L. Kubar, H. L. Treut, M. P. McCormick, G. Mégie, L. Poole, K. Powell, C. Trepte, M. A. Vaughan, and B. A. Wielicki, 2010: The CALIPSO mission: A global 3D view of aerosols and clouds. *Bull. Am. Meteorol. Soc.*, **91**(9), 1211–1229. doi:10.1175/2010BAMS3009.1.

555 Xie, S., H.-Y. Ma, J. S. Boyle, S. A. Klein, and Y. Zhang, 2012: On the correspondence between short- and long- timescale systematic errors in CAM4/CAM5 for the Years of Tropical Convection. *J. Climate*. doi:10.1175/JCLI-D-12-00134.1.

556 Zhang, Y., W. B. Rossow, A. A. Lacis, V. Oinas, and M. I. Mishchenko, 2004: Calculation of radiative fluxes from the surface to top of atmosphere based on ISCCP and other global data sets: Refinements of the radiative transfer model and input data. *J. Geophys. Res.*, **109**(D19,105). doi:10.1029/2003JD004457.

557 ——, S. A. Klein, J. Boyle, and G. G. Mace, 2010: Evaluation of tropical cloud and precipitation statistics of CAM3 using CloudSat and CALIPSO data. *J. Geophys. Res.*, **115**, D12205. doi:10.1029/2009JD012006.

00Z 15th Oct 2008	00Z 15th Jan 2009	00Z 15th Apr 2009	00Z 15th Jul 2009
06Z 16th Oct 2008	06Z 16th Jan 2009	06Z 16th Apr 2009	06Z 16th Jul 2009
12Z 17th Oct 2008	12Z 17th Jan 2009	12Z 17th Apr 2009	12Z 17th Jul 2009
18Z 18th Oct 2008	18Z 18th Jan 2009	18Z 18th Apr 2009	18Z 18th Jul 2009
00Z 20th Oct 2008	00Z 20th Jan 2009	00Z 20th Apr 2009	00Z 20th Jul 2009
06Z 21st Oct 2008	06Z 21st Jan 2009	06Z 21st Apr 2009	06Z 21st Jul 2009
12Z 22nd Oct 2008	12Z 22nd Jan 2009	12Z 22nd Apr 2009	12Z 22nd Jul 2009
18Z 23rd Oct 2008	18Z 23rd Jan 2009	18Z 23rd Apr 2009	18Z 23rd Jul 2009
00Z 25th Oct 2008	00Z 25th Jan 2009	00Z 25th Apr 2009	00Z 25th Jul 2009
06Z 26th Oct 2008	06Z 26th Jan 2009	06Z 26th Apr 2009	06Z 26th Jul 2009
12Z 27th Oct 2008	12Z 27th Jan 2009	12Z 27th Apr 2009	12Z 27th Jul 2009
18Z 28th Oct 2008	18Z 28th Jan 2009	18Z 28th Apr 2009	18Z 28th Jul 2009
00Z 30th Oct 2008	00Z 30th Jan 2009	00Z 30th Apr 2009	00Z 30th Jul 2009
06Z 31st Oct 2008	06Z 31st Jan 2009	06Z 1st May 2009	06Z 31st Jul 2009
12Z 1st Nov 2008	12Z 1st Feb 2009	12Z 2nd May 2009	12Z 1st Aug 2009
18Z 2nd Nov 2008	18Z 2nd Feb 2009	18Z 3rd May 2009	18Z 2nd Aug 2009

Table 1: List of start times for the T-AMIP2 hindcasts.

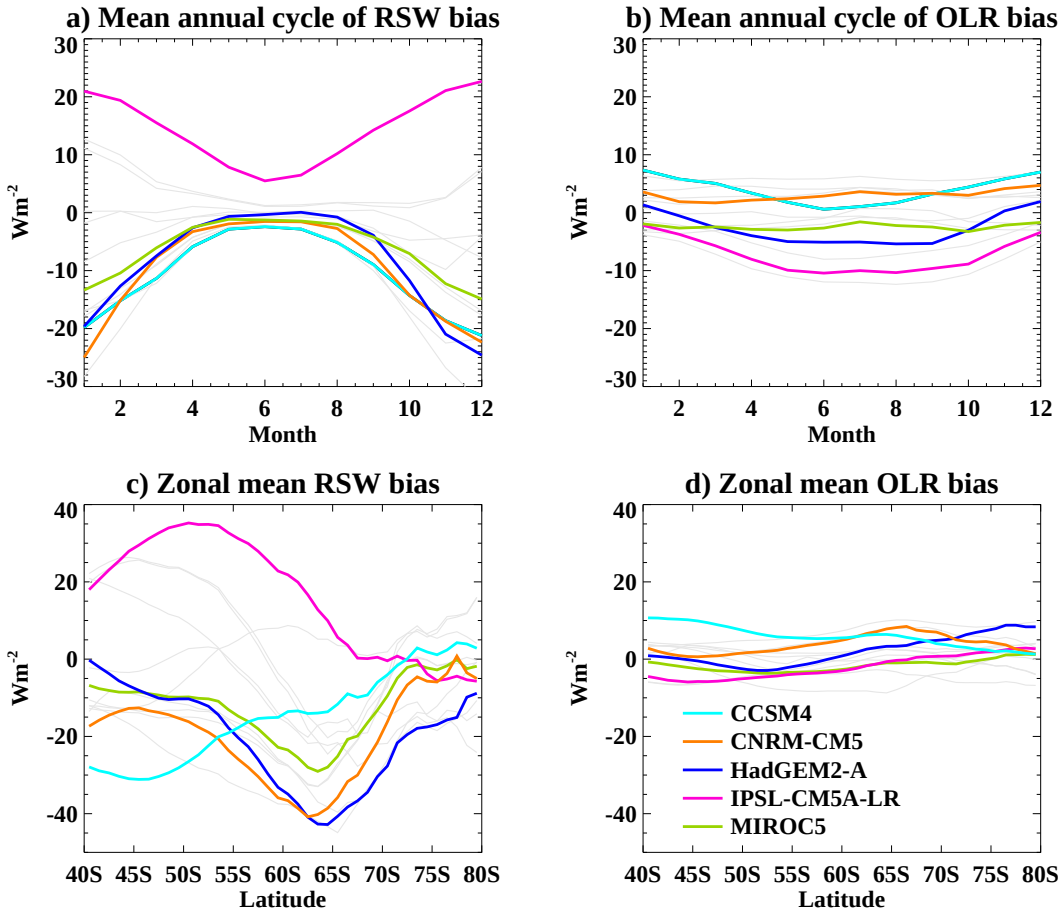

Figure 1: Mean bias in RSW and OLR in AMIP simulations compared with CERES-EBAF. a&b) show the monthly mean biases over all longitudes between 40^{o}S and 80^{o}S. c&d) show the mean DJF bias as a function of latitude. Colored lines are models which have also submitted to T-AMIP2. Gray lines show other CMIP5 AMIP models.

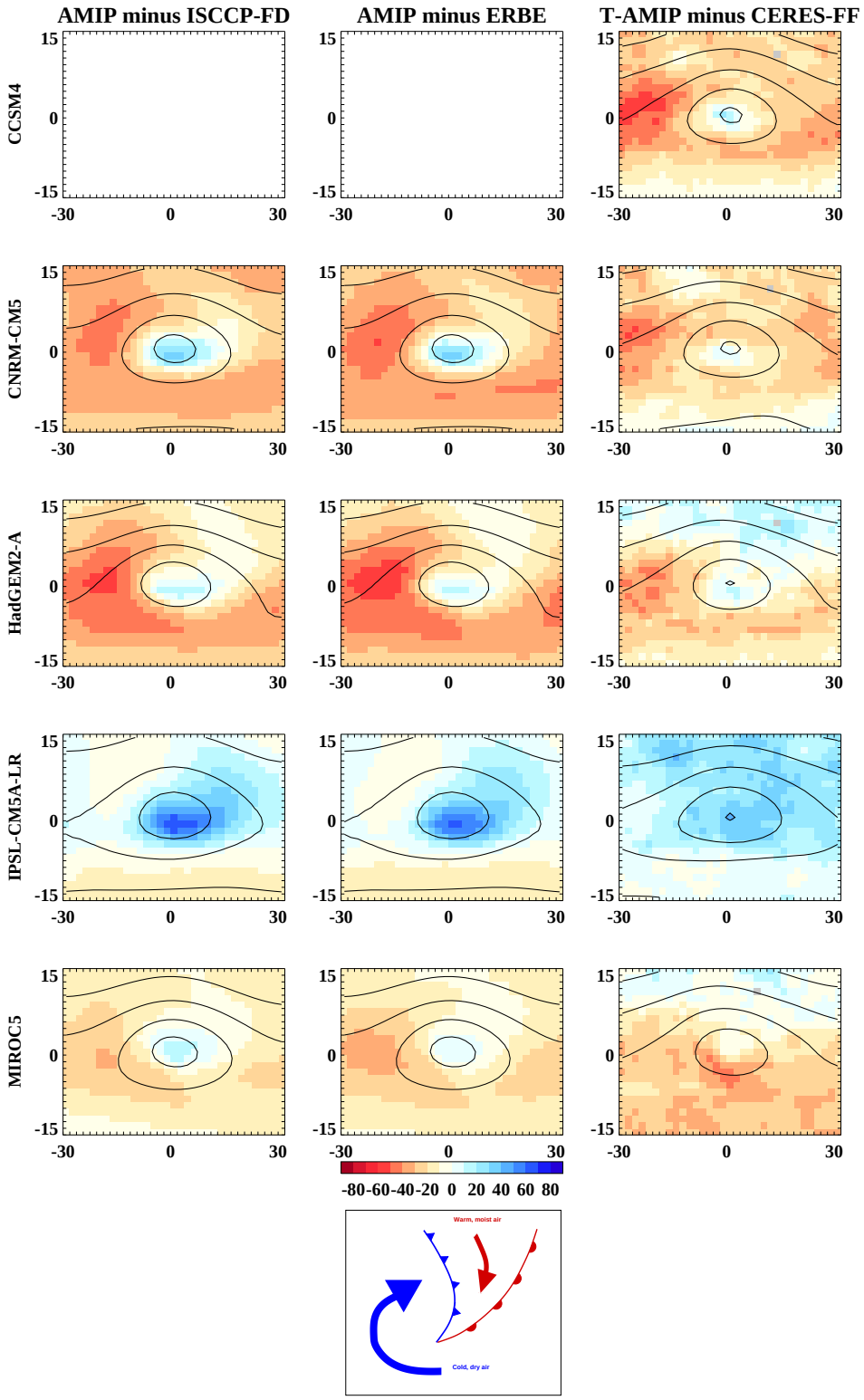

Figure 2: Mean bias in RSW for the composite cyclone over the Southern Ocean region during DJF. Contours show model mean MSLP at $8hPa$ intervals. Left column shows RSW bias in AMIP simulations against ISCCP-FD. Center column shows RSW bias in AMIP simulations against ERBE. Right column shows the day-2 RSW bias in Jan/Feb T-AMIP2 hindcasts against CERES-Flashflux. Also shown is a schematic illustrating the typical position of synoptic features in a southern hemisphere cyclone. (These diagnostics were unavailable for the CCSM4 AMIP simulation.)

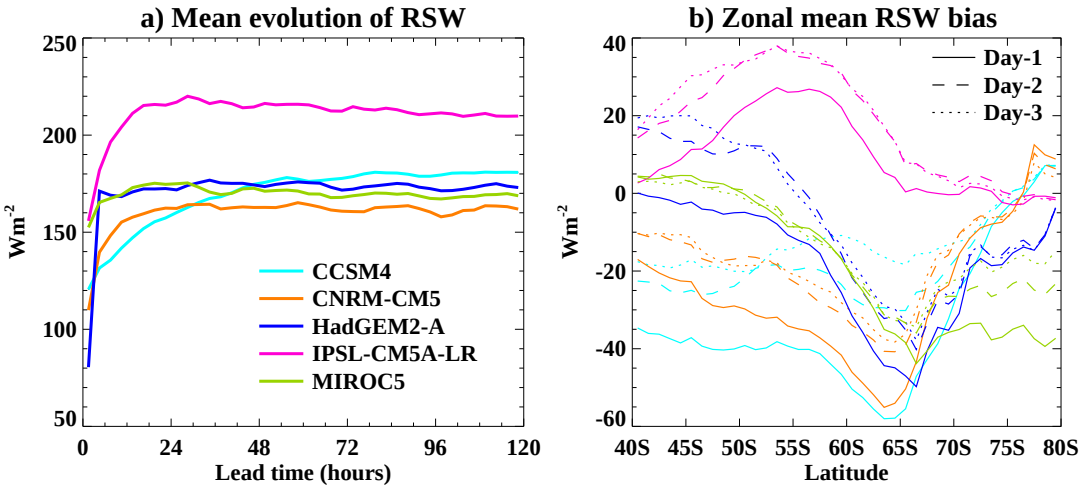

Figure 3: a) Mean evolution of RSW through the Jan/Feb 2009 T-AMIP2 hindcasts over the Southern Ocean region (55^oS – 70^oS). b) Zonal mean RSW bias against CERES-Flashflux for each of the first three days of the Jan/Feb 2009 hindcasts.

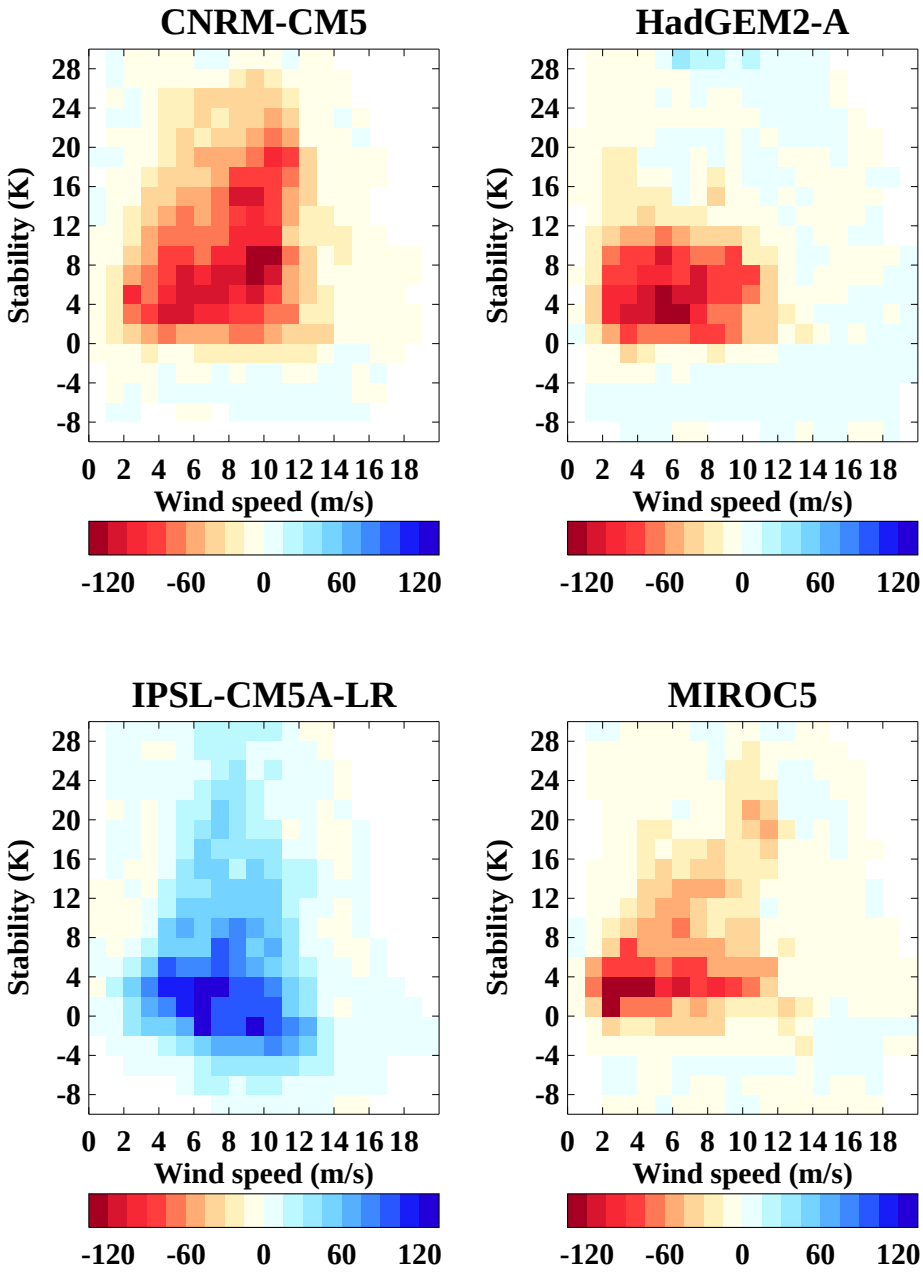

Figure 4: Histograms of mean area-weighted RSW bias against CERES-Flashflux, binned according to $10m$ wind speed and LTS. Day-2 mean values for each grid-point from the Jan/Feb hindcasts over the Southern Ocean region (55^oS – 70^oS) are used. Points beyond the limits of the axes are included within the final bin shown. (These diagnostics were unavailable for CCSM4.)

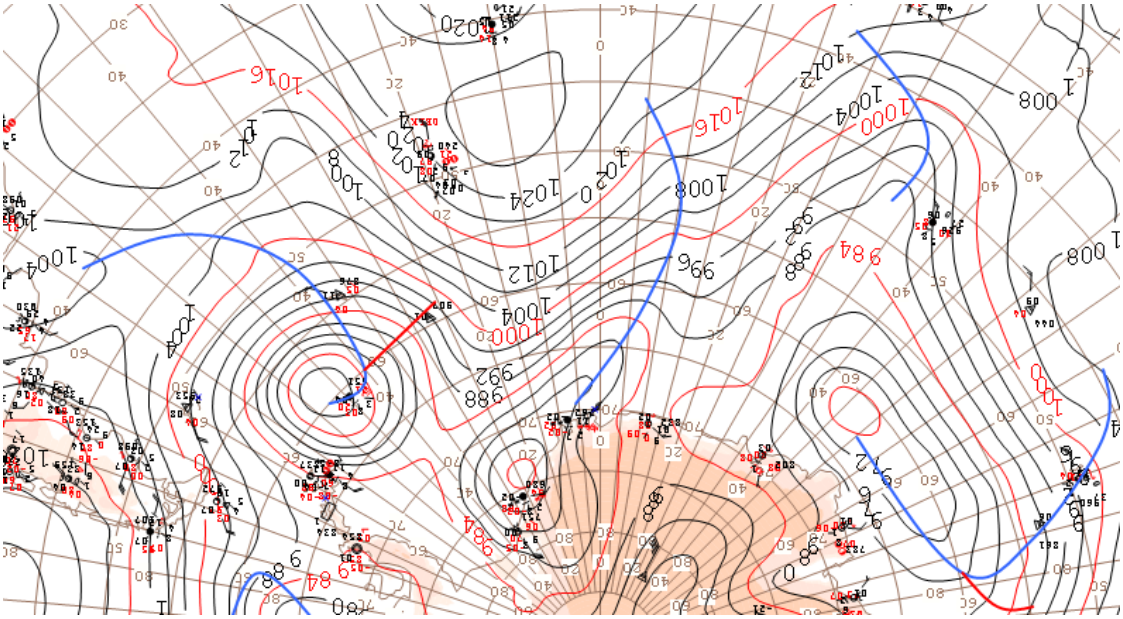

Figure 5: Met Office surface synoptic analysis for the Atlantic sector of the Southern Ocean valid 12Z 17th Jan 2009. Contours show MSLP at $4hPa$ intervals. Thicker red and blue lines mark positions of warm and cold fronts respectively.

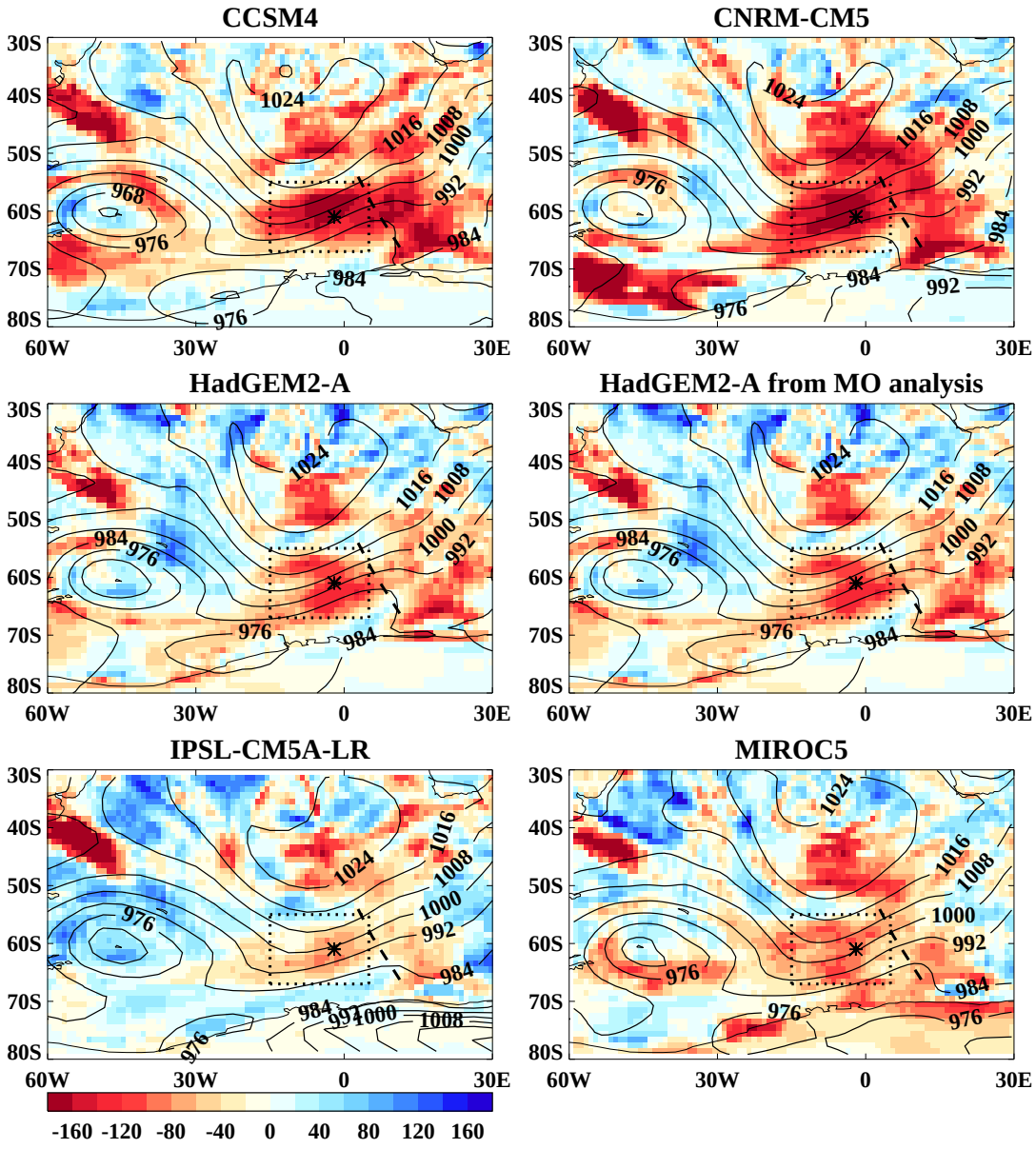

Figure 6: 24 hour mean bias in RSW against CERES-Flashflux for 17th Jan 2009 in hindcasts initialized at 06Z 16th. Contours show model forecast MSLP at 12Z 17th at $8hPa$ intervals. Dotted box and asterisk mark a region and point analyzed in this study. Dashed line marks the overpass of the satellite A-train at approximately 14:20UTC on 17th.

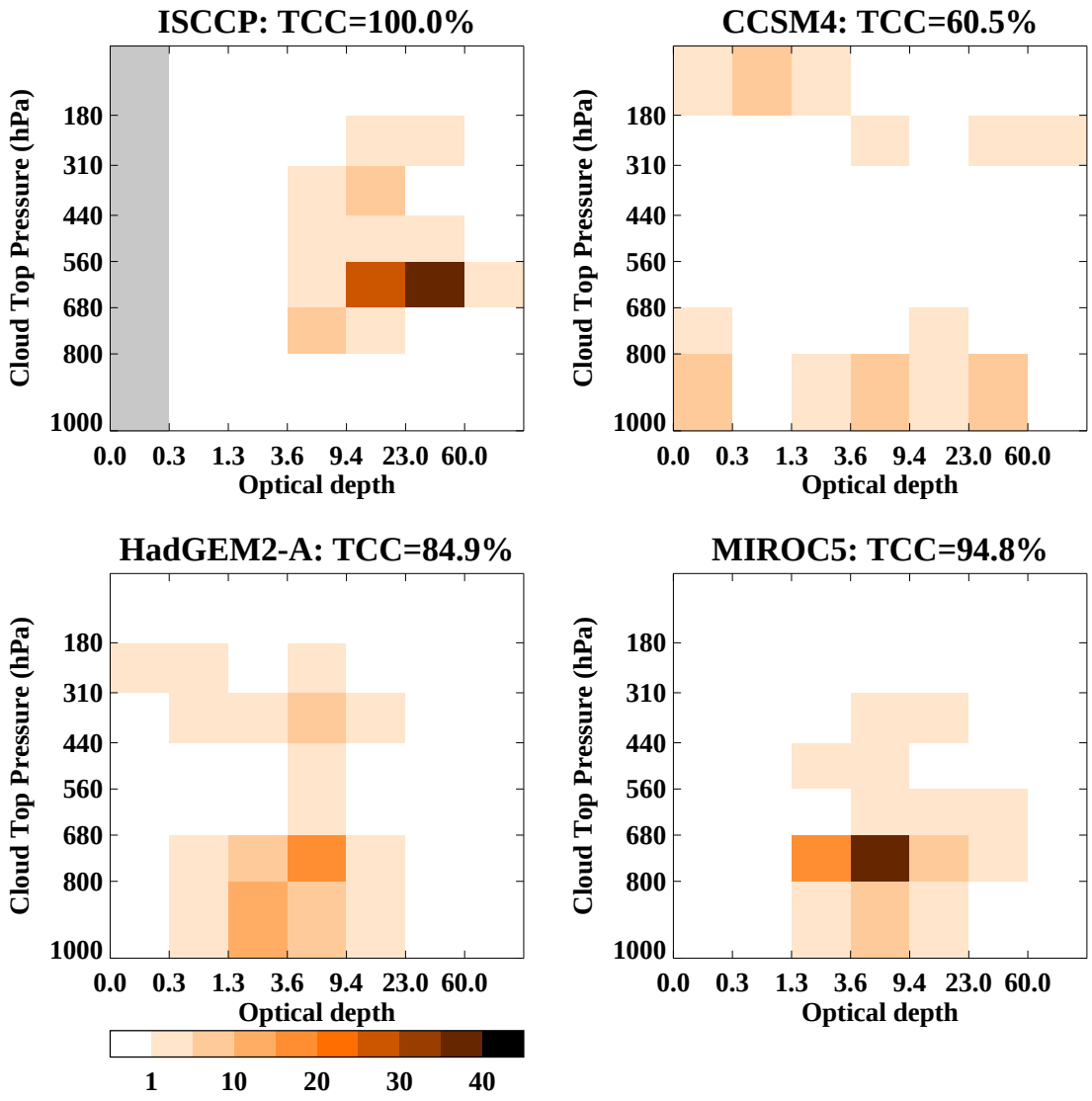

Figure 7: Histograms of cloud fraction binned according to optical depth and cloud top pressure as observed by ISCCP (top) and from the model COSP output (below) over the dotted box in Figure 6. Comparison is for 12Z 17th from hindcasts initialized 06Z 16th. The total cloud cover (TCC), which is the sum across the histogram, is given in the title of each histogram. (These diagnostics were unavailable for CNRM-CM5 and IPSL-CM5A-LR.)

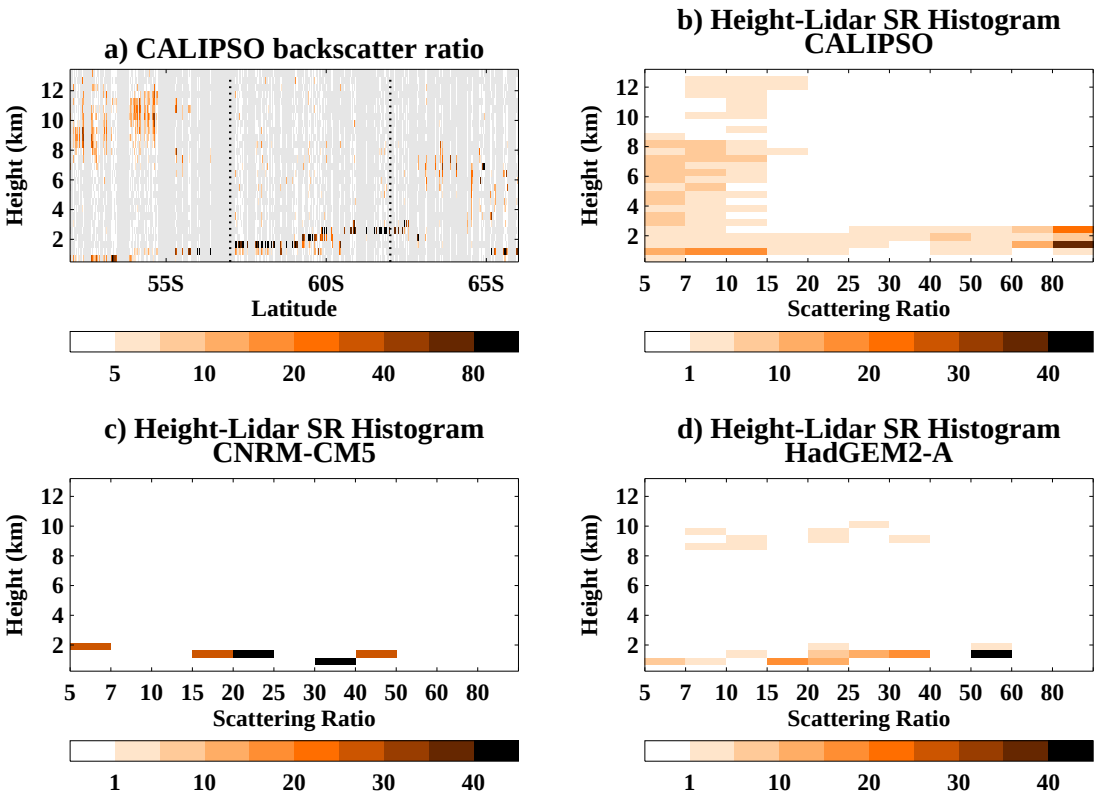

Figure 8: a) Curtain showing observed CALIPSO backscatter ratio along the transect shown dashed in Figure 6; gray columns are 'missing data'. b-d) Histogram of cloud fraction binned according to scattering ratio and altitude for the section of the transect between the dotted lines in panel a. The CALIPSO observed histogram is shown in b and model COSP output for the same section is shown in c&d. (These diagnostics were unavailable for CCSM4, IPSL-CM5A-LR and MIROC5.)

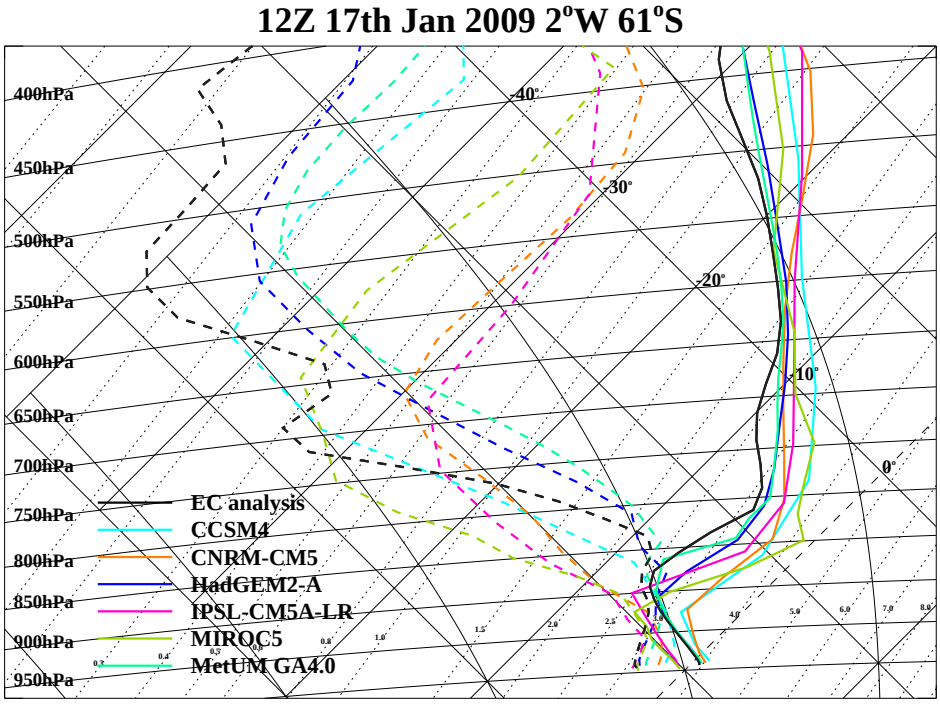

Figure 9: Standard UK tephigram showing vertical profiles of temperature (solid) and humidity (dashed) at the point marked by the asterisk in Figure 6. Profiles for the T-AMIP2 models are for 12Z 17th from hindcasts initialized 06Z 16th. Also shown is the same hindcast from a more recent configuration of the Met Office model (MetUM GA4.0), together with the ECMWF analysis profile for 12Z 17th.

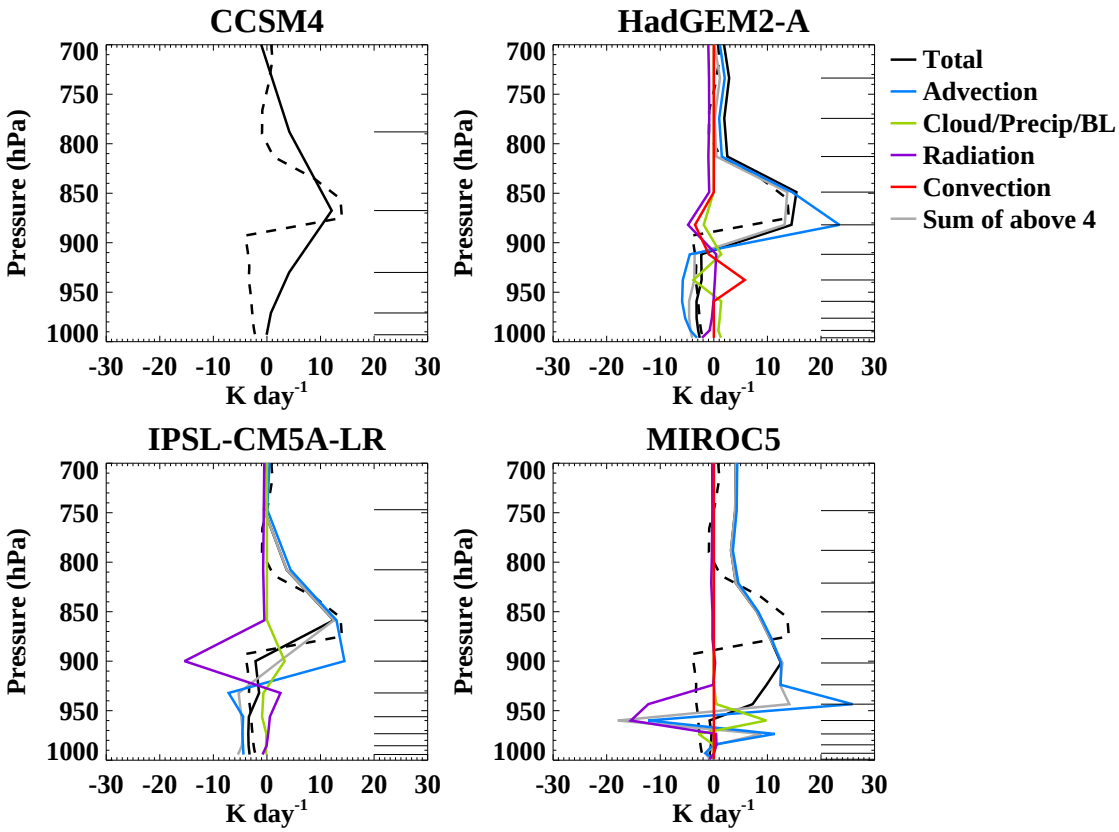

Figure 10: Mean temperature tendencies between 3 and 6 hours into the hindcast initialized at 12Z 17th Jan 2009. Total and component tendencies from various sections of the model science are shown. Dotted line shows the temperature tendency between the 12Z and 18Z ECMWF analyses. Horizontal lines on the right side of each panel mark the positions of the model levels. (These diagnostics were unavailable for CNRM-CM5. Only the total tendency was available for CCSM4 and no convective increments were available from IPSL-CM5A-LR.)

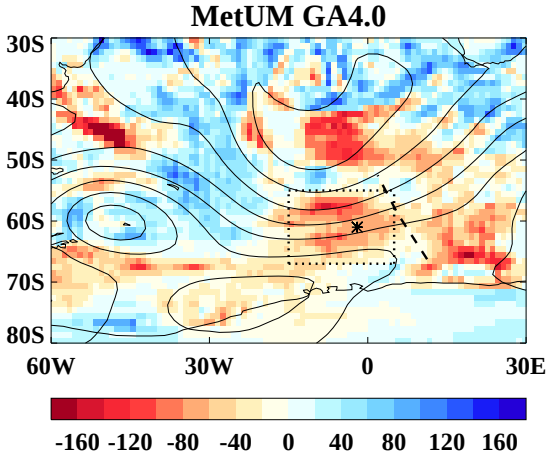

Figure 11: As Figure 6 but for MetUM GA4.0.